Response to RC1

We thank the referee for the constructive comments and thorough review for the improvement of our manuscript.

Hydrological mortality influences water availability and biomass. In this manuscript, the authors present the results of a study designed to incorporate the ELM, FATES, and Farflow to show the importance of lateral subsurface in above-ground biomass. The WTD, AGB, and Bowen ratio could be varied after using the three-dimensional hydrology model. This paper is valuable because it is the first to apply the three-dimensional hydrology model to the Earth System model. This work could be helpful for the researchers using ELM, FATES (or ED2). However, model explanation, design of experiments, and the results need more detail, and analysis is ambiguous in several places.

Thanks for the positive comments. Please see our point-by-point responses below.

Major comments

1. In section 2.1.3, while explanations of different hydraulic failure models are well described, the explanation of the biomass equation is insufficient. Clarify it (Equations calculating the above-ground biomass should be needed in this chapter).

   Above-ground biomass calculation in FATES is added and described below.

   FATES uses allometric relationships with stem diameter (*D*) to determine tree height (*h*) and crown area (*C*). There are five model options for tree height in FATES. In this study, we used a power function described in Obrien et al. (1995) for FATES input file F1 in the main text:

   $$h = 10^{(log10(D_*)\cdot a+b)} \quad \text{(R1)}$$

   $$D_* = \min\,(D, D_{max}) \quad \text{(R2)}$$

   and a Michaelis–Menten form in Martinez Cano et al. (2019) for FATES input file F2:

   $$h = \frac{cD_*^d}{k + D_*^d} \quad \text{(R3)}$$

The allometry function for crown area is

$$C = \begin{cases} fD^g & D < D_{max} \\ fD_{max}^g & D \geq D_{max} \end{cases} \tag{R4}$$

where *a, b, c, d , k, f,* and g are allometric parameters, D_{max} is diameter of plant where max height occurs.

Target biomass of leaf, structure, stem, fine root, seed, and storage are also calculated using allometry functions in FATES (Koven et al., 2020). Target biomass of fine root and storage are assumed to be linearly proportional to the target leaf biomass, and the target structure biomass is linearly proportional to the target sapwood biomass.

A power law allometric model is used for the target leaf biomass (L):

$$L = mD_*^g \tag{R5}$$

where *m* and *g* are allometric parameters, and *g* is the same as in Eq. R3.

FATES has three allometry function options to calculate target stem aboveground biomass (C_{agb}), we used the functional form in Saldarriaga et al. (1998) in F1:

$$C_{agb} = f_{agb} p_1 h^{p_2} D^{p_3} \rho^{p_4} \tag{R6}$$

and a functional form in Chave et al. (2014) in F2:

$$C_{agb} = \frac{1}{c2b} p_1 (\rho D^2 h)^{p_2} \tag{R7}$$

where f_{agb} is the fraction of stem above ground, p_1, p_2, p_3, and p_4 are allometry parameters, *c2b* is carbon to biomass ratio, ρ is the plant wood density.

Once tissue turnover and storage carbon demands are met, FATES uses a constant fraction of net primary production for seed production. Total aboveground biomass (AGB) reported in the study is the sum of leaf biomass, aboveground stem biomass and seed biomass.

2. The meteorological data which are employed in the model are from tower measurements. So, the same meteorological forcing is used in all grids? Is it a reasonable method? Moreover, which variables are used in this study (for example, temperature, precipitation, radiation).

   Note that the model can use spatial forcing, but we don't have the forcing data at the resolution of our model at BCI. We used the same meteorological forcing in all grids. It is reasonable for this site as the rain measured at the clearing near the Lutz catchment and the AVA tower (at the plateau of the 50 ha plot, ~1.25 km from the Lutz) agree pretty well on a monthly scale (Figure R1). Precipitation, air temperature, relative humidity, wind speed, surface pressure at the tower are used in this study.

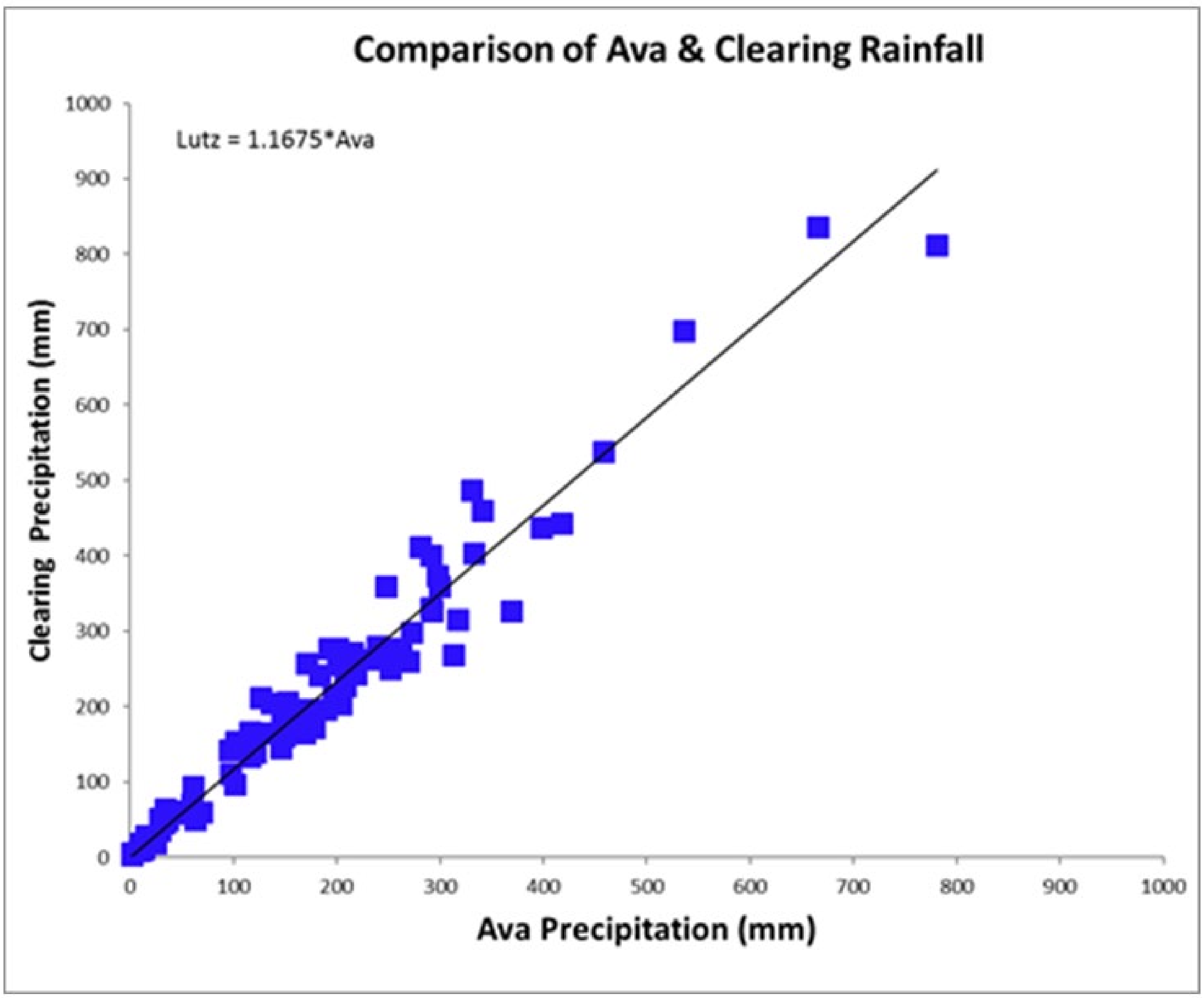

Figure R1. Comparison of precipitation at the AVA tower and at a clearing near the Lutz catchment.

This sentence now reads:

The model is driven by the same atmospheric forcing (i.e., precipitation, air temperature, relative humidity, wind speed, and surface pressure) for 2003-2016 measured at a meteorological tower near the Lutz catchment at BCI [*Faybishenko et al.,* 2018] in all grids due to the lack of spatial forcing. Comparison of the precipitation at the tower and a clearing near the Lutz catchment shows good agreement supporting the use of the same atmospheric forcing for all grids of the model.

3. Before starting the experiments, parameter tunning of the FATES model should be done because the FATES model has large uncertainties yet. The above-ground biomass in ELM-F1-M1, ELM-F1-M2, which are basic models (figure4 e) is quite underestimated because the parameter tunning was not done. After the parameter tunning was done, the results all could be different. (The effect of the 3-dimensional model on biomass may be reduced or increased.)

   The FATES parameters were selected from previous ensemble simulations for tropical forests which produced results close to the observations. We agree the results could be different after parameter tuning, but they will not affect our conclusion that water table dynamics can play large role when hydraulic failure is triggered. We made clarification in section 2.4:

   Two PFTs representing early successional and late successional species are simulated at the same time in competition with each other using two input files of plant traits selected from previous ensemble simulations that best matched observations for tropical forests [Chen et al. 2022, Huang et al. 2020]. Further parameter tuning is out of the scope of this work.

4. With the 3-dimensional model, the AGB and GPP were simulated bigger at lowland than at highland. Do you have observation or reference? Does the fact that the simulated wetter soil in lowland using a three-dimensional model show the improving model performance or just model comparison?

   Unfortunately, we don’t have spatial observations that cover both lowland and highland to validate the simulated AGB and GPP. The difference at lowland and highland is model comparison. It is clarified as:

   Note this is based on model comparisons. Spatial observations at those locations are needed to validate the model but such observations are not currently available.

5. In section 3.2, while the results of ELM-PF-FL-M1 are well written, the explanation of ELM-PF-F2-M1 was quietly insufficient. Moreover, what is difference between F1 and F2? This explanation could be added in section 2.4.

   Thanks for the suggestion. We added explanation of ELM-PF-F2-M1. Please see our response to the following Comment 6. The main difference between F1 and F2 are the allometric models described in Comment 1 above. F2 also has lower carbon starvation mortality rate. These two input files are now included in the Supplement. The following descriptions are added in section 2.4:

   The allometric models for tree height and target stem aboveground biomass in F1 are defined in Eqs. 4 and 9, respectively, and those in F2 are defined in Eqs. 6 and 10, respectively.

   The complete parameters for F1 and F2 are included in the Supplement.

6. With F2, the AGB increases, but GPP decreases. In general, the AGB is positive correlated with GPP, but it is not in this case. This could be discussed with a trait of plants. (Line 476). And it would be better to think about it the direct impact of plant trait on AGB and the indirect impact (plant trait -> WTD -> AGB)

   As the soil moisture simulated using F1 and F2 are close, GPP is mainly affected by growth allometry while AGB is the result of both growth and mortality. From Fig. R2 we can see that using F2 results in larger growth rates and significantly lower mortality rates, thus increasing AGB for F2 compared to F1. However, simulation with F2 results in much lower exposed leaf area index, thus lower GPP compared to that with F1. If we use the same input file, increasing only V_{cmax} will increase both AGB and GPP as shown in Fig. 5. The following are added in the manuscript in response to Comments 5 and 6. Fig. S1 below refers to Fig. R2 in this response.

   The spatial standard deviations (STDs) of the aforementioned-variables of interest for simulation ELM-PF-F2-M1 are slightly smaller than for ELM-PF-F1-M1. The difference in STD between ELM-PF-F2-M1 and ELM-PF-F1-M1 is larger for VWC, LH, and SH compared to GPP, AGB and WTD. With this plant traits F2, AGB increases by 47.5% and GPP decreases by 19% on average (Fig. 4). As the soil moisture (VWC) simulated using ELM-PF-F1-M1 and ELM-PF-F2-M1 are close , GPP is mainly affected by growth allometry while AGB is the result of both growth and mortality. Using plant traits F2 results in larger growth rates and significantly lower mortality rates (Fig. S1), thus increasing AGB for F2 compared to F1. However, simulation with F2 results in much lower exposed leaf area index, thus lower GPP compared to that with F1.

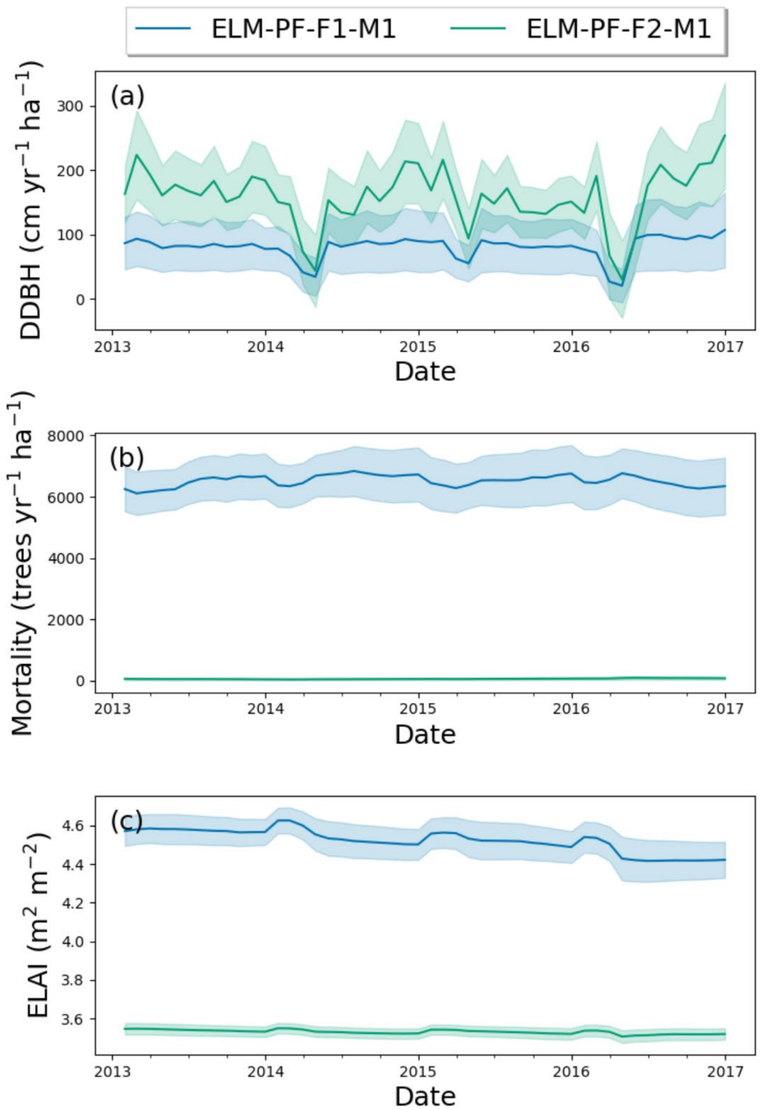

Figure R2. Growth (increment of diameter at breadth height, DDBH) (a) vs. mortality (b), and exposed leaf area index (c) for ELM-PF-F1-M1 and ELM-PF-F2-M1. Solid line is spatial average and shaded area is the standard deviation over the simulation domain.

7. In-Line 479~484, I was confused that the experiment with adjusting the VCMAX was shown suddenly. There is no experiment in Table 1 and in Section 2.4. And why did you multiply 1.9? Is there any reference?

   As it was pointed out in Comment 3, our untuned parameters underestimated AGB. This experiment was introduced to show that tuning parameters can improve the simulated AGB, but the AGB variability is still too small compared to the observation. The factor of 1.9 is determined through trial and error, and there is no reference. We removed the experiment from the text and Fig. 5 to avoid confusion. It now reads

AGB can be further increased by parameter tuning for better agreement with observations, but we don't expect it to significantly change the AGB variability.

8. In section 3.3. it is very interesting about the response of AGB, VWC, Water table depth, H mortality and C mortality to M1, M2 and M3 model. In Line 514, you mentioned that M2 simulates wetter soil in the dry season compared to M3. But it seems likely that simulated AGB in M3 is more than M2 (Figure 6 (a), (b)). Could you add the results of AGB in M2 and M3 in dry seasons and discuss that why is AGB in M3 is more than M2? To my knowledge, the vegetation in wetter soil makes more carbon, especially in dry seasons.

   Thanks for the comment. The higher AGB in M3 is mainly at locations near the shallow water table where M3 simulates lower hydraulic mortality. We added the results and discussion shown below. Fig. S2 in the following refers to Fig. R3 in this response.

   AGB from ELM-PF-F1-M2 is smaller compared to that from ELM-PF-F1-M3, especially at locations where the water table is shallow (WTD < 2 m). That is due to the higher mortality rate triggered by hydraulic failure in ELM-PF-F1-M2 at those locations (Fig. 6c), resulting in fewer grids with AGB > 4.5 kg C/m^2 (Fig. S2).

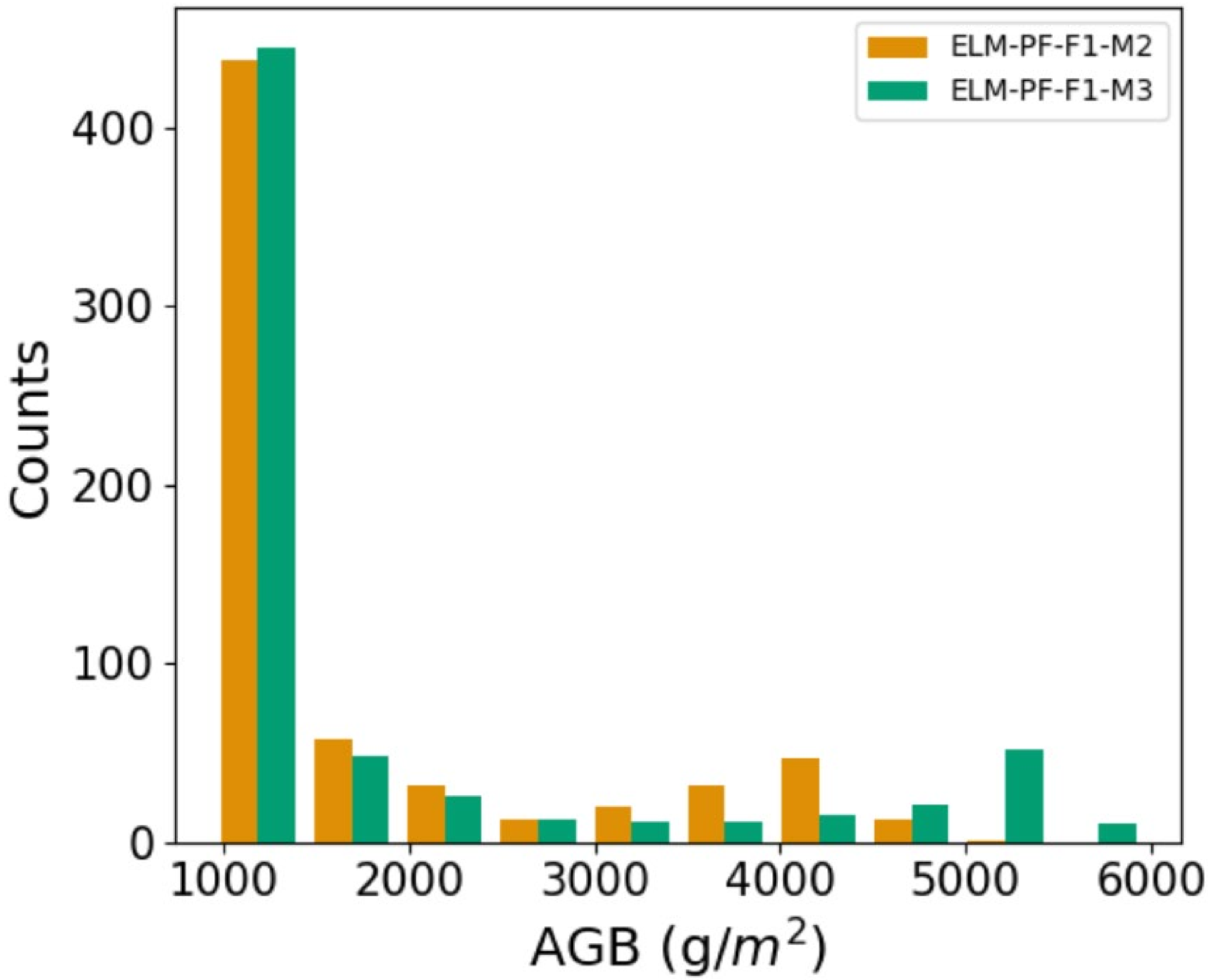

Figure R3. Histogram of simulated AGB for ELM-PF-F1-M2 and ELM-PF-F1-M3

9. In section 3.4, the description of variable importance of AGB, WTD was written. How did you calculate the importance? It should be added in section 2.5.

   We added the following in section 2.5:

   To calculate the permutation importance, a reference score (prediction error) for a trained regression model *m* is first calculated. Each feature *j* (a column) in the training or testing dataset is randomly shuffled to generate a corrupted dataset and the score of the model *m* on the corrupted dataset is calculated. The shuffling and corrupted dataset score computation are repeated multiple times. The importance of feature *j* is computed as the difference between the reference score and the arithmetic mean of the scores of the model *m* on the corrupted datasets. This is documented in https://scikit-learn.org/stable/about.html#citing-scikit-learn.

10. It is necessary to first show the performance of how well AGB and WTD are predicted by the random forest method. It is necessary to show the importance after presenting the results and discussing them. Moreover, why did you select elevation, slope, convexity, and vwc using the random forest? Is this sufficient to explain the importance of AGB and WTD?

    Thanks for the comment. We adjusted the order of results and the importance and it now reads:

    The RF models have shown good performance. They can explain 90% and more of the variance (VAR_{ex} in Table 2) in AGB and WTD for both the training data and the unseen test data, suggesting the predictors selected are sufficient to explain AGB and WTD. They perform better for AGB than for WTD with mean absolute percentage error (MAPE) less than 10% as opposed to 30% for WTD (Table 2). All explanatory variables used as predictors in the RF models can capture portions of the variability of the simulated AGB and WTD, but the relative importance of the predictors is different for the different ELM-PF models (Fig. 7).

    The reason we selected elevation, slope, convexity, and vwc is because they have previously been shown to affect AGB and WTD, and the reason was provided at the beginning of section 2.5 in the original submission. Our selection of the predictor variables is supported by the results of our random forest models showing that those predictors are sufficient to explain AGB and WTD.

11. This paper has novelty because a three-dimensional model is incorporated with Earth system modeling. In this paper, the domain is very small, while the domain and spatial resolution of research using the Earth system model are still large. Using a three-dimensional model may have inaccurate results or no significant impact with large resolution. It would be good to add a discussion on using a three-dimensional model in a global scale.

    Thanks for the positive comment and suggestion. We added the following discussion:

    Using a three-dimensional model in current Earth system models typically run at ~100 km grid resolution may yield inaccurate results or have no significant impact on vegetation dynamics. A reasonable grid resolution for groundwater flow simulation is around 1 km (Xie et al. 2020 and references therein). Moving from 100 km to 1 km resolution for global scale vegetation dynamics simulation coupled with a 3D integrated hydrologic model is computationally challenging at present, but it may be a realistic goal with advances of computation power and architecture in the future. The model in this study provides opportunities for improving hydrological, ecological, and meteorological predictions of Earth system models.

Minor comments

Contracted words are inappropriate in a science paper.

Line 103 during1983 -> during 1983

Thanks for the correction.

Line 198 The words of leaf area index are repeated

Thanks. Deleted.

Line 293~295 The sentence about ET could be deleted

The sentence is rewritten as to describe the source of the observations. It now reads

Observed evapotranspiration (ET), gross primary production (GPP), sensible heat flux (SH), and latent heat flux (LH) at the site were obtained from an eddy-covariance system installed in July 2012 on the AVA tower.

Line 318~319 and 321~322 These sentences should be rearranged (-> in section 2.3)

Thanks. They are moved to section 2.3.

Line 333 2)->1)

Thanks. Corrected.

Line 349~357 The explanations about spin up and experimental design are quietly hard to understanding. The case 3 was run 200 years and after that additionality runs for 16 years was done due to the comparing case 6 and case 7. It would be nice to show this in Table 1. And did you run the CASE 5 after case 4?

Yes, Case 5 was run after Case 4. Table 1 is revised as in the following:

Table 1. Definition of model experiments with ELM, PF, F, and M denoting E3SM land model, ParFlow, different parameters for plant traits, and different mortality models, respectively. “K” in experiment name of Case 5 indicates soil property derived from Kupers et al. [2019b] is used. Extra 16 years of simulation were conducted for four experiments.

Cases	Model Experiments	Plant Traits	Soil Property	ParFlow	Drought Mortality Model	Extra simulation years for model comparison
1	ELM-F1-M1	F1	S1	No	Eq. (11)	0
2	ELM-F2-M1	F2	S1	No	Eq. (11)	0
3	ELM-PF-F1-M1	F1	S1	Yes	Eq. (11)	16 from Case 3
4	ELM-PF-F2-M1	F2	S1	Yes	Eq. (11)	0
5	ELM-PF-F2-M1,K	F2	S2	Yes	Eq. (11)	16 from Case 4
6	ELM-PF-F1-M2	F1	S1	Yes	Eq. (13)	16 from Case 3
7	ELM-PF-F1-M3	F1	S1	Yes	Eq. (14)	16 form Case 3

VWC is already shown in Figure 6, but there is no description of VWC in section 3.3. The first introduction of VWC is in section 3.4

Thanks! We added the introduction of VWC after “soil moisture content” in section 3.3.

Line 610 adds the discussion why the results of test data was not good.

We discussed that “the data is sparse and/or the observed AGB may depend on other factors such as soil heterogeneity and nutrient availability” at the end of the paragraph in the original submission.

References:

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B. C., Duque, A., Eid, T., Fearnside, P. M., Goodman, R. C., Henry, M., Martínez-Yrízar, A., Mugasha, W. A., Muller-Landau, H. C., Mencuccini, M., Nelson, B. W., Ngomanda, A., Nogueira, E. M., Ortiz-Malavassi, E., Pélissier, R., Ploton, P., Ryan, C. M., Saldarriaga, J. G., and Vieilledent, G.: Improved allometric models to estimate the aboveground biomass of tropical trees, Glob. Change Biol., 20, 3177–3190, 2014

Martínez Cano, I., Muller-Landau, H. C., Wright, S. J., Bohlman, S. A., and Pacala, S. W.: Tropical tree height and crown allometries for the Barro Colorado Nature Monument, Panama: a comparison of alternative hierarchical models incorporating interspecific variation in relation to life history traits, Biogeosciences, 16, 847–862, https://doi.org/10.5194/bg-16-847-2019, 2019.

O'Brien, S. T., Hubbell, S. P., Spiro, P., Condit, R., and Foster, R. B.: Diameter, Height, Crown, and Age Relationship in Eight Neotropical Tree Species, Ecology, 76, 1926–1939, 1995.

Saldarriaga, J. G., West, D. C., Tharp, M. L., and Uhl, C.: Long-Term Chronosequence of Forest Succession in the Upper Rio Negro of Colombia and Venezuela, J. Ecol., 76, 938–958, 1988.

Xie, Z., Wang, L., Wang, Y., Liu, B., Li, R., Xie, J., et al. (2020). Land surface model CAS-LSM: Model description and evaluation. Journal of Advances in Modeling Earth Systems,12, e2020MS002339. https://doi.org/10.1029/2020MS002339